

# News Analysis and Summarization from RSS feeds

Supratik Chatterjee  
B.Tech, CSE, Third Year,  
SRM IST,  
Vadapalani, Chennai,  
Tamil Nadu, India.  
RA1511003040196

Shubham Bagadiya  
B.Tech, CSE, Third Year,  
SRM IST,  
Vadapalani, Chennai,  
Tamil Nadu, India.  
RA1511003040258

Sanjeev Siva  
B.Tech, CSE, Third Year,  
SRM IST,  
Vadapalani, Chennai,  
Tamil Nadu, India.  
RA1511003040193

***Abstract*—In the modern day field of media, news is majorly affected by the social platforms such as Facebook, Twitter, etc. It becomes necessary to understand the facts of an event, and to quantify or relate them to understand it. Most humans do not have the dedication or time to study all the events at all the time, as time is always limited. However, an aggregation and analysis system on a computer could do it for them.**

***Index Terms*—Natural Language Processing, Web Scraping, Text Mining, news, Co-occurrence Analysis, Information Centralization, Machine Learning, Clustering, News Aggregation**

## I. INTRODUCTION

The aim of the project is to centralize information collected from news websites for connected and summarized reading. This aims to remove co-occurrence of facts, and find discrepancies of figures, so that one might deal with them as an outlier. The rest of the data must be catalogued and indexed such that they may be referenced easily. When this is hooked up with a Machine Learning tool, should be able to predict outcomes, provided the dataset given to it has been collected for a duration long enough for it to work on developing itself.

Collection of information has to be from the several RSS feeds found online. Using those as the data point one can fetch the information and centralize the resources for further analysis and prediction. This project will be an exercise in centrality, distributed computing, natural language processing, data mining, clustering and machine learning.

## II. LITERATURE REVIEW

There exists no such tool that analyzes and aggregates the feeds for others. However, there are applications made to organize the feeds for easy viewing. There are several problems for such an application. These problems areas follows :

- 1) Articles are repeated
- 2) No summarization or focus on facts
- 3) No check on the feeds
- 4) No interrelation between points of interest
- 5) Hard to look up